

Journal: *Biocatalysis and Biotransformations*

Biodiversity for biocatalysis: a review of the α/β -hydrolase fold superfamily of esterases-lipases discovered in metagenomes

MANUEL FERRER^{1*}, RAFAEL BARGIELA¹, MÓNICA MARTÍNEZ-MARTÍNEZ¹,
JAUME MIR², RAINHARD KOCH³, OLGA V. GOLYSHINA⁴ & PETER N. GOLYSHIN⁴

¹*Institute of Catalysis, Consejo Superior de Investigaciones Científicas (CSIC), 28049 Madrid, Spain.*

²*Biochemize SL, Barcelona Advanced Industry Park, BCNactiva, 08042 Barcelona, Spain*

³*Bayer Technology Services GmbH, 51373 Leverkusen, Germany*

⁴*School of Biological Sciences, Bangor University, Gwynedd LL57 2UW, UK*

*Correspondence: Institute of Catalysis, Consejo Superior de Investigaciones Científicas (CSIC), 28049 Madrid, Spain. Email: mferrer@icp.csic.es

Short title: Biodiversity for biocatalysis

Keywords: Biocatalysis, biotransformation, esterases, extremophiles, lipases, metagenome

Abstract

Natural biodiversity undoubtedly inspires biocatalysis research and innovation. Biotransformations of interest also inspire the search for appropriate biocatalysts in nature. Indeed, natural genetic resources have been found to support the hydrolysis and synthesis of not only common but also unusual synthetic scaffolds. The emerging tool of metagenomics has the advantage of allowing straightforward identification of activity directly applicable as biocatalysis. However, new enzymes must not only have outstanding properties in terms of performance but also other properties superior to those of well-established commercial preparations in order to successfully replace the latter. Esterases (EST) and lipases (LIP) from the α/β -hydrolase fold superfamily are among the enzymes primarily used in biocatalysis. Accordingly, they have been extensively examined with metagenomics. Here we provided an updated (October 2015) overview of sequence and functional datasets of 288 EST–LIP enzymes with validated functions that have been isolated in metagenomes and (mostly partially) characterized. Through sequence, biochemical and reactivity analyses we attempted to understand the phenomenon of variability and versatility within this group of enzymes and to implement this knowledge to identify sequences encoding EST–LIP which may be useful for biocatalysis. We found that the diversity of described EST–LIP polypeptides was not dominated by a particular type of protein or highly similar clusters of proteins but rather by diverse non-redundant sequences. Purified EST–LIP exhibited a wide temperature activity range of 10–85°C, although a preferred bias for a mesophilic temperature range (35–40°C) was observed. At least 60% of the total characterized metagenomics-derived EST–LIP showed outstanding properties in terms of stability (solvent tolerance) and reactivity (selectivity and substrate profile), which are the features of interest in biocatalysis. We hope that, in the future, the search for and utilization of sequences similar to those already encoded and characterized EST–LIP enzymes

from metagenomes may be of interest for promoting unresolved biotransformations in the chemical industry. Some examples are discussed in this review.

Introduction

α/β -Hydrolase fold superfamily of esterases (EST; EC 3.1.1.1) and lipases (LIP; EC 3.1.1.3) have received considerable attention in recent decades for two primary reasons. First, they are widely distributed and widespread in nature within microbial communities operating in most environments where they have important physiological functions (at least one can be found in each bacterial genome). Second, they belong to one of the most important groups of biocatalysts for chemical synthesis with multiple commercial preparations available for industrial use (Nagarajan 2012; Turner & Truppo 2013). Indeed, most screening requests at the industrial scale for chemical synthesis address aldo-keto reductases (KREDs) followed by transaminases (TAs) and lipases.

Esterases and lipases have been identified on the basis of their substrate specificity (Bornscheuer et al. 2002), which can be extensively modified using protein engineering approaches (Jochens et al. 2011). EST enzymes hydrolyze solutions of water-soluble short acyl chain esters and are mostly inactive against water-insoluble long chain tri-acyl-glycerols, which, in turn, are specifically hydrolyzed by LIP. From the comparison of structural features of sequence-related EST and LIP, it appears that LIP, unlike EST, commonly displays a significant difference in the distribution of hydrophobic amino acid residues in the vicinity of their active site and the existence of a particular surface domain that specifically interacts with lipid-water interfaces. Indeed, clear evidence of interfacial activation was given by the first lipase structure in open and closed conformations, namely human pancreatic lipase (Winkler et al. 1990) and *Rhizomucor miehei* lipase (Brady et al. 1990). Since then, different structures of esterases and lipases have corroborated the differences in interfacial activation and substrate specificities (Kazlauskas & Bornscheuer 1998). Despite the existence of examples that contradict this rule (Anobom et al. 2014), EST and LIP have been found to support the hydrolysis and synthesis of a wide range of similar compounds, including esters and amides (Turner & Truppo 2013;

Anobom et al. 2014) for which the molecular basis was, in some cases, established (Suplatov et al. 2012; Lan et al. 2015). Promiscuous behavior has also been reported, such as the unusual ability to hydrolyze carbon-oxygen bonds in a broad spectrum of esters as well as carbon-carbon bonds in aromatic ring fission products that have been demonstrated for a set of EST-LIP from the α/β -hydrolase fold superfamily (Alcaide et al. 2013). A promiscuous lipase (CalB from *Candida antarctica*) was found not only to support ester hydrolysis but also to support fast carbon-carbon bond formation with the addition of 1,3-dicarbonyls to α/β -unsaturated carbonyl compounds (Svedendahl et al. 2005). In both cases, single point mutations at the proximity of the active site were responsible for this promiscuity. These examples have *inter alia* demonstrated that esterases/lipases are even more versatile than previously thought and that promiscuity may offer advantages for a number of biochemical applications in media containing organic solvents (Bornscheuer & Kazlauskas 2004; Hult & Berglund 2007; Kapoor & Gupta 2012; Sharma & Kanwar 2014). In addition to the differences in protein fold and specific mutations, the versatility of the serine-histidine-aspartate catalytic triad typical of EST-LIP (and many other enzymes from the α/β -hydrolase fold superfamily) may play a role in defining promiscuity events (Rauwerdink & Kazlauskas 2015).

Based on the comparison of primary structures, 14 families of sequence-related EST and LIP have been found. Indeed, LIP and EST families were grouped into Family I (covers the most abundant true LIP), Family II (also called the GDSL family), Family III, Family IV (also called the HSL family), Family V, Family VI, Family VII and Family VIII (Arpigny & Jaeger 1999). Subsequent studies led to the discovery of new enzymes that could not be grouped in the existing 8 families according to the Arpigny & Jaeger (1999) classification. The 9th family of bacterial esterases (Family IX) was created on the basis of the depolymerase PhaZ7 (Handrick et al. 2001). The hyper-thermophilic esterase EstD from *Thermotoga maritima* was categorized into Family X. Metagenomics-derived lipolytic enzymes LipG (Lee et al. 2006) and LipEH166

(Kim et al. 2009) were the 11th and 12th bacterial esterase family members. Family XIII was created after the discovery of esterases from *Bacillus* (Ewis et al. 2004; Liu et al. 2004; Rao et al. 2013). The most recently described bacterial lipolytic family was Family XIV (Rao et al. 2013).

Cultivated microorganisms represent a very small fraction of the actual diversity of microorganisms present in nature (Chistoserdova et al. 2014; Kyrpides et al. 2014), and therefore, the former provide a very limited picture of the actual capability of natural niches and the microbes and enzymes they contain to convert chemicals of interest. Metagenomics can potentially provide access to all genetic resources present in an environment, regardless of whether or not they belong to microorganisms that can be cultured in the laboratory. Thus, metagenomics represents an alternative approach to unraveling the biodiversity of microorganisms and activities (Kyrpides et al. 2014; Yarza et al. 2014). Functional metagenomic analyses, which are based on the detection of a phenotypic change in the bacterial host due to the acquisition of a gene coding for a particular activity, have the advantage of allowing the straightforward identification of activities even if their coding genes are so unrelated to those in the databases that they cannot be detected by sequence analysis (Ferrer et al. 2015).

Due to their importance in industrial chemical transformations, EST–LIP enzymes are important biocatalysts as shown by the high number of scientific publications and patents for technological forecasting (Daiha Kde et al. 2015). Thus, they have been extensively examined by metagenomics approaches. Functional screenings based on the development of a color resulting from multiple substrate cleavage have been successfully used to detect genes coding for EST–LIP activity in a high number of metagenomic functional studies (Reyes-Duarte et al. 2012; Ferrer et al. 2015). However, while the intrinsic biodiversity at the sequence level was expected to be very high, a large fraction of the identified EST–LIP enzymes from uncultured

microorganisms were uncharacterized in terms of their potential for biocatalysis and biotransformations, as we will discuss later. In addition, the scale of their applications in industrial biocatalysis was rather disappointing (Ferrer et al. 2015). Indeed, the extraordinary potential of environmental enzymes has not yet been sufficiently exploited, mostly due to our limited knowledge of their biochemistry and performance. In addition, the fact that the market is dominated by highly versatile commercially available preparations such as Novozym CALB or Lipolase (marketed by Novozymes S/A), which are used for the production of fine chemicals (Daiha Kde et al. 2015), makes it difficult to introduce new enzymes to the market in case these contenders do not possess outstanding properties and/or cannot be produced easily at a reasonable cost. By 2015, a bibliographic search retrieved 1671 patents and 2223 scientific publications reporting esterase/lipase applications in different fields (Daiha Kde et al. 2015). The reason that these enzymes dominate the applications (e.g., patents) are their promiscuity, robustness, solvent tolerance and, most importantly, scalability, which means that large amounts can be produced at low costs. Noticeably, only a few examples of applications of EST–LIP from metagenomes have been patented, e.g., US8580549B2, US20090078384A1, WO2013073760A1, WO2014147486A2, CN104328132A, EP04015920.4, WO2007017181. Therefore, EST–LIP enzymes discovered with metagenomic analysis still represent a tiny fraction of the patented and applied enzymes for technological forecasting.

A better knowledge of the characteristics of enzymes directly extracted from the environment by metagenomic approaches would improve our ability to select those with high potential as biocatalysis (Bommarius 2015). So far, there is no single resource that provides a compilation of metagenome protein sequence information and biochemical and reactivity datasets with information regarding substrate specificity and solvent and thermal tolerance for esterases/lipases. We assumed that such analysis would be a valuable tool for academic and industrial researchers. In an attempt to compile this information, we have provided updated

datasets regarding EST–LIP discovered by metagenomics approaches. We performed a comparative sequence and functional analysis of these datasets. The identified metagenomics EST–LIP allowed examination of them in the context of diversity and divergence and their potential for biocatalysis. A large number of sequences were found to encode, higher than anticipated to some extent, EST–LIP with relatively high stability in media containing organic solvents and broad and unusual specificity in terms of substrate and catalytic promiscuity. Those enzyme candidates may be of interest to support relevant biotransformations in the chemical industry, some of which are herein summarized. Note that we are aware that companies such as Diversa Corporation in USA (which no longer exists) and BRAIN in Germany have also discovered and patented numerous EST–LIP from metagenomes. For example, in an extensive study, 350 novel lipases and esterases discovered from environmental DNA samples were characterized for their capacity to modify lipids (Bertram et al. 2008). However, sequences encoding such hydrolases were not considered in the present review due to the limited information available.

Diversity of EST–LIP genes

All metagenomic-derived EST–LIP-like sequences were obtained from the National Center for Biotechnology Information (NCBI) with accession numbers indicated in Supplementary Table I. We restricted the search for sequences encoding EST–LIP with validated functions. Under these conditions, a total of 288 sequences of valid EST–LIP members were retrieved (for details see Supplementary Table I). Note that additional sequences (from genome origin) encoding EST–LIP members have been compiled recently and are available as a database (Kourist et al. 2010); therefore, they were not considered in this study.

A multiple alignment of MetaGenomic protein Sequences (referred to as MGS) was used to build a phylogenetic tree using the neighbor-joining algorithm. Model Genome protein

Sequences unambiguously ascribed to each of the 14 different Families of the α/β -hydrolase fold superfamily of esterases-lipases (referred to as GS-FI to GS-FXIV), were also included. Most of the metagenomic EST-LIP sequences were clearly ascribed to existing families (Figure 1), with representatives in all of the 14 distinct families known to date. Note that, as mentioned in the Introduction section, the 11th and 12th bacterial esterase family members were only covered by metagenomic derived lipolytic enzymes LipG (Lee et al. 2006) and LipEH166 (Kim et al. 2009).

Pairwise amino acid sequence identity analysis suggests a large divergence at the sequence level. Indeed, identity values for all 288 sequences ranged from 0 to 98.6%, with most of the enzymes sharing only 10-15% identity (Figure 2). This suggests that each of the sites screened for esterase/lipase activity contain distinct enzyme arsenals and that actual metagenomic screen programs have covered a high diversity of EST-LIP genes. This is most likely due to selection in the different habitats of communities that host phylogenetically distinct bacterial species. Supplementary Table I provides information about which habitats have been more abundantly explored by metagenomics approaches for EST-LIP. Briefly, soil was the major source of EST-LIP (98 in total), followed by marine habitats (87), microbiota associated with multiple hosts (45), freshwater like habitats (22), compost (14), and wastewater, activated sludge or biogas slurry (8). Their sequences were also compared with the sequences available in the NCBI non-redundant public database. Protein sequences were on average 68% similar to homologous proteins in the database. This reflects the huge yet undiscovered fraction of EST-LIP from microbial species populating the environment and the completely different evolutionary origins of enzymes from source (uncultured) microorganisms.

Taken together, the phylogenetic analysis of the metagenomic sequences reported to date revealed a highest diversity of EST-LIP and showed that, in spite of the relatively large number

of EST–LIP sequences already known, we may still be far from having complete knowledge of the biodiversity of these enzymes in nature.

Hydrolytic activities of EST–LIP under different conditions

Enzymes are only used for industrial processes if they comply with a number of criteria. These include: *i*) the ability to maintain activity and stability under harsh reaction conditions and a broad range of temperatures (enzymes should at least be stable at room temperatures for a significant period of time - storage might be an additional issue, which can refer to detergent enzymes applied in warmer countries); and *ii*) to be active and stable under a broad range of water-deficient reaction conditions and under process conditions (e.g., active for 12-24 h) (Viader-Galván et al. 2010; Spickermann et al. 2014; Zuhse et al. 2015). Accordingly, the optimal temperature for activity and stability in, and resistance against, media containing organic solvents, are important parameters to examine. Note, however, that the scalability and production of both the esterase/lipase and the synthetic product obtained are also important parameters that are not commonly evaluated; for example, many EST–LIP proteins are difficult to express (Troeschel et al. 2012).

The application of EST and LIP from solvent-tolerant microbes has recently become an attractive resource for biocatalysis (Gaur & Khare 2011), as they may contain organic solvent-tolerant enzymes. In addition, a link between tolerance to high salt concentrations, which is common in halo-tolerant microorganisms, and solvent-resistance has been extensively confirmed (de Lourdes Moreno et al. 2013). For this reason, within metagenomic-derived EST–LIP with confirmed functions, we screened for those that have been reported to be stable or were activated in the presence of salts (e.g., NaCl) and solvents (e.g., methanol [MetOH], ethanol, butanol, isopropanol, dimethylsulfoxide [DMSO], dimethylformamide [DMF], and acetonitrile [ACN], to name some) to identify those that have the potential to be used in

processes under low-water conditions. Note that the performance in the presence of salts and solvents was not tested for 210 out of 288 of the EST–LIP enzymes (Figure 3). Within those for which data are available (78), a total of 47 were solvent-tolerant (see blue accession numbers in Figure 1). This represents approximately 60% of the total characterized EST–LIP from metagenomes. A total of 27 enzymes (35%) were additionally found to be salt-tolerant (see blue and red accession numbers in Figure 1). Although their activity and stability in the presence of solvents were not tested, they would most likely be resistant to solvents according to the link between salt and solvent-resistance (de Lourdes Moreno et al. 2013); however, experimental evidence is needed to verify this theory and to identify which solvents these enzymes are tolerant of. The high percentage of solvent- or salt-tolerant EST–LIP enzymes, together with the fact that many of these enzymes were isolated from habitats characterized by low salinity (with no potential halophiles; Supplementary Table I), suggests that solvent-resistance is a general characteristic of EST–LIP enzymes independent of their phylogenetic origin.

Prediction of solvent tolerance based on sequence information requires a special discussion. In this context, many studies have dissected the features of proteins tolerant to salts/solvents, revealing that they usually contain a large excess of acidic amino acid residues compared to those of non-halophilic organisms (Frolow et al. 1996; Madern et al. 2000). The fact that a high proportion of the characterized EST–LIP enzymes from metagenomes were solvent/salt-tolerant, prompted us to examine the occurrence of charged amino acids in the sequences of EST–LIP proteins. For this purpose, all 288 sequences were used. We observed a preponderance of acidic residues (independent of the microbial origin of the sequences), with most proteins having isoelectric point (pI) of sequences as low as 4.0-4.5 (Figure 4). This agrees with the fact that the majority of EST–LIP discovered by metagenomics approaches were potentially solvent-tolerant. Note however, that few proteins that were solvent-tolerant (AEH57833.1, AFK83589.1, EU660533.1, AKG92633.1, AFS34517.1, ADN26553.1 and KP861228) were characterized by

pI from 8.2 to 11.3 (Supplementary Table I), suggesting that additional factors play a role in solvent resistance. These proteins were active and stable in the presence of 15% (v/v) methanol, butanol or DMSO. However, a careful inspection of the data reported for other metagenomic EST-LIP with lowest pI values revealed they are the most active in the presence of higher concentrations (> 30% v/v) of above and other organic solvents that included isopropanol (GQ426329), DMF (ADB22436.1, ADG03645) and ethanol (ABY60416). This suggests that higher solvent resistance is a characteristic mostly associated with sequences that have lower pI values. It is plausible that new sequences encoding EST-LIP phylogenetically similar or proximal to those reported as being highly solvent-tolerant (see Figure 1 in blue) may exhibit similar properties.

One important question regarding solvent resistance is the nature of the solvents used in process development in a chemical company, so they can be further tested to find industrially-relevant EST-LIP in metagenomes. In this context, the most important solvents are MetOH and ACN. Other solvents, which can be mixed with water to a certain extent are used, including acetone, isopropanol and, less often, 1,4-dioxan. Solvents that are used on a technical scale, but not mixable with water, are methyl-t-butylether (MTBE), dichlormethane, toluol, methyl isobutyl ketone (MIBK) or 2-methyl tetrahydrofurane. Specifically, DMF and dimethylacetamide (DMA) are most often avoided on the technical or commercial.

In terms of optimal temperature (T_{opt}) for activity, purified EST-LIP exhibited a wide optimal temperature range of 10-85°C, although a preferred bias for a mesophilic temperature range (35-40°C) was found (Figure 5). This is consistent with the fact that the majority of environments that have been subjected to naïve activity screens are characterized by moderate temperatures (Ferrer et al. 2015). Indeed, only 10 EST-LIP have been isolated and characterized from thermal vent or hot spring-like habitats (Supplementary Table I). Note that the differences in T_{opt} are independent of the phylogenetic proximity of their sequences (Figure 1), as enzymes belonging

to the same sub-families can show different temperature profiles and vice versa (Supplementary Table I). Although we also observed that moderately low temperature environments contain microbes with EST-LIP enzymes that are mostly active at temperatures as high as 70°C (Alcaide et al. 2015a); this is the case of the MGS-M2 esterase (KF831415.1; Protein Data Bank code 4Q3L) isolated from the hyper-saline deep sea lake *Medee* that never exists at a temperature higher than 16.5°C (Alcaide et al. 2015a). Unknown structural factors were responsible for such behavior, which may, most likely, be a consequence of protein adaptation to multiple extreme environmental factors, particularly high salinity and high pressure, which are characteristic of some deep sea environments within the Mediterranean Sea (Alcaide et al. 2015a). Therefore, organisms that grow best at moderate temperatures may contain EST-LIP enzymes that are active at elevated temperatures because they have a more stable protein structure. Conversely, it has been reported recently that proteins from cold-adapted organisms inhabiting Polar (and deep sea) waters that rarely are warmer than 4-6°C exhibit generally higher temperature optima (T_{opt} from 20 to 50°C) (Kube et al. 2013).

Considering thermal extremes, a number of relevant features were found. Within thermophilic members of EST-LIP we observed an enzyme (AB302136) that is capable of hydrolyzing poly-(DL-lactic acid) polyesters at 70°C and also a solvent-tolerant hydrolase (AFS34517.1) that is highly specific for (*R*)-ibuprofen-phenyl ester and is also capable to synthesize 1-propyl laurate and 1-tetradecyl myristate at 70°C. Additionally, we observed a solvent-tolerant enzyme (4P6B_A) highly specific for (*S*)-ketoprofen at 65°C. Of particular interest is an EST enzyme (MGS-HA1; KF831418) from a superficial seawater sample at a hydrothermal vent at Saint Paul Island that is most active at 70°C, possesses an ample hydrolytic performance, and is able to hydrolyze at least 33 distinct esters, including bromo, chloro, fluoro and iodo alkyl and aryl esters, tri-O-acetyl-glucal, α -D-glucose penta-acetate, and hydroxycinnamic esters, and is highly specific for a number of chiral esters (Alcaide et al. 2015a). Within cold-active EST-LIP

enzymes we found an enzyme (KC438416) that can hydrolyze dialkyl phthalates to their corresponding monoalkyl phthalate esters at 10°C, an enzyme (KC986400) that can hydrolyze polyester substrates at 15 °C, including poly-(DL-lactic acid), and two enzymes (ABY61092.1 and AEH57832.1) that are capable of preferentially hydrolyzing (*S*)-racemic ofloxacin butyl ester at 20°C. An enzyme from an evaporite karstic lake (LAE5; SRA059294) that is most active at 12°C was also capable of hydrolyzing 14 structurally distinct esters.

Taken together, within the herein reviewed EST–LIP enzymes from metagenomes, a large fraction include enzymes that have properties of interest for biocatalysis, such as tolerance to solvents, wide ranges of temperatures and ample substrate specificity.

Reactivity of EST–LIP from metagenomes with different substrates

Other important parameters defining the relevance of EST–LIP in biocatalysis are the substrate profile, high stereo-selectivity and high turnover rate (Singh et al. 2010). Enzymes should display novel and broad substrate/product ranges (Ferrer et al. 2015) and exquisite regio-, enantio- and stereo-selectivity properties. In the particular case of EST–LIP, the following features are appreciated: *i*) the ability to transform large substrates, the larger the better; *ii*) the ability to cleave only one chiral ester when offered a racemic mixture of two esters; and *iii*) the capacity to perform the reverse action, e.g., synthesis of ester and amide bonds; etc. EST–LIP that have these properties can have potential applications in biocatalysis (Steele et al. 2009). In addition, two criteria have become increasingly important, namely, the possibility to produce the enzyme in large quantities in a fermenter, the ability to use the enzyme in a bioreactor and the amenability of an enzyme property of interest to enzyme evolution. If an enzyme has a flexible backbone, an expressible enzyme that fits the same process conditions, has solvent tolerance, but different substrate specificity can be developed; it is anticipated that promiscuous EST–LIP enzymes are perfect starting points for this application (Ferrer et al. 2015).

Synthetic activities

Reactivity characteristics have been reported for 235 out of the 288 EST–LIP-like enzymes isolated to date by metagenomic approaches. The majority of them were tested in terms of hydrolytic capacities, whereas synthetic capabilities were tested and verified for only 5 of them. This last set included a purified lipase (AY304500) from oil-contaminated soil that had a conversion degree of biodiesel production of nearly 92% in a trans-esterification reaction using olive oil and methanol (Zheng et al. 2013). Additionally, the first example of a metagenome-derived lipase (LAE6; SRA059294) capable of producing sucrose esters that have broad applications (Plou et al. 2002) has been reported; the enzyme, isolated from the metagenome of an evaporite karstic lake allowed trans-esterification reactions to produce sucrose esters by using long chain vinyl esters (Martínez-Martínez et al. 2013). In addition, the trans-esterification of glyceryl triacetate with methanol (EU660533.1; Meilleur et al. 2009) and the synthesis of 1-propyl laurate and 1-tetradecyl myristate (AFS34517.1; Chow et al. 2012) were successfully reported. More recently, using immobilized preparations of a lipase (KM023399) from soil samples contaminated with fat, enantio-selective conversions have been achieved for the trans-esterification of the alcohol (*R,S*)-1-phenylethanol with vinyl acetate (Alnoch et al. 2015); an enantiomeric excess (*ee*) higher than 95% for the *R*-isomer was achieved.

Hydrolytic activities

Regarding hydrolytic capacity, the majority of EST–LIP from metagenomes were tested only for their ability to hydrolyze common esters, including tri-acyl-glycerol and *p*-nitro-phenyl esters, with acyl chains ranging from acetate to palmitate. The available data are consistent with the highest esterase over lipase character of the identified and characterized enzymes, as suggested by the preponderance of enzymes acting over shorter (<C₁₀) esters. Therefore, true lipases are sparsely identified by metagenomics approaches. Based on substrate profiling, and

considering substrates other than tri-acyl-glycerol and *p*-nitro-phenyl esters, validated functions and activities were established. Out of the 235 characterized enzymes, 6 hydrolases (isolated from seawater) could be classified as having the unusual ability to hydrolyze carbon-oxygen bonds in a broad spectrum of esters as well as carbon-carbon bonds in aromatic ring fission products (Alcaide et al. 2013). Two were cutinases (AEV21261.1 and 4EB0) found in compost (Mayumi et al. 2008; Sulaiman et al. 2012) with poly(ϵ -caprolactone)- and/or polyethylene terephthalate (PET)-degrading activity; thus, they may have potential applicability for surface modification and degradation of PET and polymers. Sixteen had the capacity to hydrolyze esters in a component present in plant-like polymers. They include 5 acetyl-xylan esterases (CAJ19128.1, CAJ19130.1, CAJ19122.1, CAJ19113.1 and CAJ19109.1) (Ferrer et al. 2005), and 11 feruloyl esterases with the ability to hydrolyze to different extend esters of hydroxycinnamic acids that include methyl sinapate, methyl ferulate, ethyl ferulate, methyl *p*-coumarate and methyl caffeate (see Supplementary Table I). Two of them (AEI54552.1, AEI54551.1) additionally release ferulic acid and diferulic acid from untreated crude plant cell wall materials and one (AKA87410.1) releases ferulic acid from rice bran, wheat bran, wheat-insoluble arabinoxylan, corn fiber, switchgrass, and corn bran. One (KC438416) was an esterase able to hydrolyze dipropyl phthalate, dibutyl phthalate, and dipentyl phthalate to their corresponding monoalkyl phthalate esters. Two (KF709432, KF705200) had the capacity to hydrolyze phytate and can be thus considered to be phytate esterases. Four (KF801579, KC986400, KC986401, AB302136) were efficient for hydrolyzing poly-(DL-lactic acid) (PLA) polyesters and are therefore PLA depolymerases. One (HQ147564) was shown to remove 80% of tannins from a green tea infusion on the first treatment after immobilization and is thus a tannase. Two (EU285670.1, ABN58716.1) were reported to be phospholipases, with EU285670.1 being also capable of hydrolyzing triglycerides and phosphatidylcholine. A total of 6 enzymes (AFU54388.1, AFK83589.1, ACH88047.1, AEA07655.1, AEA07653.1, and AGT17593.1) had the capacity to hydrolyze not only esters but also β -lactam antibiotics, and

thus they are esterases with β -lactamase activity. Of particular interest is a β -lactamase (ACH88047.1), which, together with the capacity to hydrolyze short-to-medium chain esters, specifically hydrolyzed esters of tertiary alcohols such as linalyl acetate (3,7-dimethyl-1,6-octadien-3-yl acetate) and had detectable promiscuous β -lactam hydrolytic activity. The rest of the characterized enzymes could be unambiguously classified as being common esterases/lipases with different hydrolytic capacities and substrates specificities (see below).

In terms of enantio-selective conversions, a number of relevant enzymes with outstanding reactivity were found. For example, the stereo-selective conversion (>91% *ee*) of esters of ibuprofen have been achieved using lipases/esterases from oil contaminated soil (AFS34517.1; ABI94371.1). Stereo-selective conversion for solketal esters (98% *ee*) has been achieved using lipases/esterases from oil deep-sea hypersaline anoxic basins (Ferrer et al. 2005). While a novel (*S*)-ketoprofen-specific esterase (4P6B_A) has been identified using pools of metagenomes from the Microbank of Microbial Genomics and Application Center (Taejon, South Korea) (Yoon et al. 2007), (*R*)-ketoprofen-specific esterase (AAY45707.1) was found in soil metagenomes (Kim et al. 2006). AAY45707.1 efficiently hydrolyzed (*R,S*)-ketoprofen ethyl ester with slight enantio-selectivity toward (*R*)-ketoprofen ethyl ester, while 4P6B_A mostly hydrolyzed (*S*)-ketoprofen. Twenty-one metagenome derived esterases have also been effective for the enantio-selective kinetic resolution of phenylalkyl carboxylic acids (56 to >99% *ee*) and the kinetic resolution of 1,1,1-trifluoro-2-phenylbut-3-yn-2-yl acetate and 3,7-dimethyl-1,6-octadien-3-yl acetate (enantiomeric ratio [*E*] > 100); the kinetic resolutions were performed in preparative scales (Kourist et al. 2007; Fernández-Álvaro et al. 2010). Enzymes AEH57833.1 and AEH57832.1 hydrolyzed methyl 3-phenylglycidate, a chiral synthon for the synthesis of the Taxol® side chain. Using (*R,S*)-1-phenylethyl acetate, Im-LipG9 (KM023399) showed excellent enantio-selectivity for the *R*-isomer (*E* > 200), giving an *ee* of higher than 95% for the products at 49% conversion. Using a racemic mixture, the capacity to hydrolyze preferentially (*S*)-

racemic ofloxacin butyl ester with an *ee* of 70.3% was found using an esterase (EU195806) from arctic sediment. Two novel esterases (AAY90130.1 and AAZ48934.1) isolated from drinking water also showed high enantio-selectivity for (+/-)-1-octin-3-ol, *R*-(+)-3-chlor-1-phenyl-1-propanol, trimethylsilylbutinol, *cis/trans*-1,2-cyclohexanediol and isopropylidenglycerol acetate. Finally, an esterase (AAY90130.1) was highly enantio-selective for (+)-menthylacetate. Three esterases from Lake Arreo (LAE1, LAE3 and LAE4; SRA059294) showed for (*R,S*)-methyl mandelate, whereas one lipase showed enantio-selectivity (LAE6; SRA059294) for (*R,S*)-glycidyl butyrate. *E* values varied from 816 (for LAE3) to 15 (for LAE6), 8.3 (for LAE1) and 2.0 (for LAE4). LAE1 and LAE3 did show enantio-preference for methyl-(*R*)-(-)mandelate, whereas LAE4 for methyl-(*S*)-(+) mandelate; LAE6 preferred (*R*)-(-)-glycidyl butyrate. LAE6 further exhibited activity toward γ -butyrolactone, with (*S*)-(+)-glycidyl butyrate being the best enantiomer. Three esterases (KF831416/MGS-RG1; KF831417/MGS-RG2; and KC986402/MGS-RG3) of gill chamber-associated microbiota in the deep-sea shrimp *Rimicaris exoculata* were also found to be enantio-selective, to different degrees, for 5 chiral esters. Based on a calculation of *E* values for separate enantiomers, the enantiomeric ratios and substrate preferences were as follows: *i*) 158.4 ± 11.0 (preference for methyl-(*S*)-mandelate), 19.6 ± 0.8 (for (*S*)-glycidyl-4-nitrobenzoate) and 20.0 ± 0.7 (for (*S*)-methyl-3-bromo-2-methyl propionate) for MGS-RG1; *ii*) 16.1 ± 1.3 (for methyl-(*S*)-lactate), 16.3 ± 1.1 (for (*S*)-glycidyl-4-nitrobenzoate), 26.1 ± 2.2 (for (*S*)-methyl-3-bromo-2-methyl propionate) and 65.4 ± 7.9 (for methyl-(*S*)-mandelate) for MGS-RG2; and *iii*) 300 ± 9.3 for MGS-RG3 and menthyl-(*R*)-acetate. LIPESV12_9 (KR919661), LIPESV12_24 (KP861227), LIPESV12_26 (KP861228) from hydrothermal vent sediments of the Levante Bay (Placido et al. 2015) were also found to be enantio-selective, to different degrees and preferences, for at least eight chiral esters, including methyl-(\pm)-mandelate, methyl-(\pm)-lactate, (\pm)-menthylacetate, (\pm)-neomenthyl acetate, (\pm)-glycidyl 4-nitrobenzoate, (\pm)-pantolactone, and (\pm)-methyl -3-hydroxybutyrate.

Other features of interest were also found, such as the ability of the esterase ACJ07038.1 to degrade pyrethroid pesticides and rho-nitrophenyl esters of medium-short chain fatty acids. Additionally, an esterase (AAY90130.1) shows a high level of activity against a wide range of substrates including one secondary ester, 7-[3-octylcarboxy-(3-hydroxy-3-methyl-butyloxy)]-coumarin, which is normally un-reactive. A lipase (AAZ67909.1) supports the hydrolysis of the triglyceride derivative 1,2-di-O-lauryl-rac-glycero-3-glutaric acid 6'-methylresorufin ester. Hydrolysis of aryl esters such as methyl phenyl-acetate and phenyl acetate have also been reported (KM042178; and enzymes reported in Martínez-Martínez et al. 2013, Alcaide et al. 2015a, and Placido et al. 2015). The esterase AEL88620.1 catalyzed the deacetylation of 1- and 3- acetyl and 1,3-diacetyl esters; the enzyme reactivated chloramphenicol from its acetyl derivative by counteracting the chloramphenicol acetyltransferase activity in *Escherichia coli*. A lipase (ABI94371.1) demonstrated selectivity for esters of primary alcohols, whereas esters of secondary or tertiary alcohols remained almost unconverted.

Few recent examples examined enzyme reactivity using an ample set of structurally different esters (> 140), therefore not only the esterase/lipase character but also the substrate and reaction promiscuity and enantio-selectivity features were evaluated. Recent examples include the following: seven esterases-lipases (SRA059294) from the metagenome of Lake Arreo, an evaporite karstic lake in Spain (Martínez-Martínez et al. 2013), 3 esterases of gill chamber-associated microbiota (KF831416, KF831417 and KC986402) in the deep-sea shrimp *Rimicaris exoculata* (Alcaide et al. 2015b), five esterases (KF831414, KF831415, KF831419 – KF831421) from five distinct deep-sea (3040-4908 m depth) and moderately warm biotopes (Alcaide et al. 2015a), one esterase (MGS-HA1; KF831418) from a marine hydrothermal habitat (Alcaide et al. 2015a), 5 hydrolases able to hydrolyze carbon-carbon and carbon-oxygen bonds from marine habitats (Alcaide et al. 2013), 3 esterases (KR919661, KP861227, and KP861228) from hydrothermal vent sediments (Placido et al. 2015), and 5 cold-active and salt-resistant

esterases (JX133669, JX133670, JX133672-JX133675) from marine metagenomes (Tchigvintsev et al. 2015). These studies demonstrated that esterases/lipases discovered in metagenomes show reactivity and promiscuity levels that are superior to those of the homologous hydrolases reported to date. This is exemplified by the capacity of a single EST to hydrolyze C-C and C-O bonds (Alcaide et al. 2013), or the ability of a single protein to hydrolyze tri-acyl-glycerols, (non)halogenated alkyl and aryl esters, cinnamoyl and carbohydrate esters, lactones, and chiral epoxides to a similar extent (Martínez-Martínez et al. 2013).

The fact that at least 10% of the total EST–LIP described to date in metagenomes do have reactivity features of interest in biocatalysis (see above) together with the fact that some of them show activity towards chemicals that are normally un-reactive (see accession numbers marked with a black circle Figure 1; outstanding enzymes are summarized in Table I), clearly support metagenomics as a relevant tool for the discovery of biocatalysts. The isolation of new enzyme variants with which one could approach the synthesis of chemical intermediates of interest may be thus relevant for biocatalysis. Although substitutions at a restricted set of amino acids have been shown to extensively influence enzyme properties (Alcaide et al. 2013), it is generally accepted that proteins sharing a high level of homology may have similar biochemical characteristics. The information herein summarized can be used for the future prediction and identification of key gene sequences encoding EST–LIP enzymes with relevant physical-chemical performance and reactivity. In brief, each query sequence from a known metagenome that matches a given protein sequence of those experimentally validated enzymes with outstanding properties (see details in Supplementary Table I and Figure 1) may be selected as having potential application for biocatalysis. For this reason web-based resources for filtering predicted open reading frames in the metagenomic DNA sequences for similar EST–LIP should

be of interest; these sequences can be subjected to intensive gene synthesis and further protein characterization.

Performance of EST–LIP from metagenomes compared to commercial preparations

The new esterases/lipases to be discovered by metagenomics (and other) approaches, including the ones reported to date and those that are herein summarized, should have not only outstanding properties (see above) but also have properties superior to those of other well established commercial lipases such as CalB from *Candida antarctica*; this lipase is one of the most successful commercially available lipase preparations (Novozymes A/S, Bagsvaerd, Denmark). Few examples in the literature revealed the potential of metagenomics to identify esterases/lipases with substrate profiles and performances similar or even superior to CalB. Thus, by examining the hydrolytic capacity of a set of 101 different esters, it was found that the lipase LAE6 (SRA059294) from the metagenome of an evaporite karstic lake showed a reactivity profile similar to that of CalB (Martínez-Martínez et al. 2013): it was active and stable at 60°C. Additionally, a novel lipase (AFS34517.1) obtained from metagenomic libraries constructed from enrichment cultures was able to hydrolyze a number of industry-relevant substrates with an *ee* of 99% as well as the ability to synthesize 1-propyl laurate and 1-tetradecyl myristate at 70°C, with rates similar to CalB (Chow et al. 2012). Furthermore, a novel enzyme (MPlaG) from a tidal flat metagenome (EP2784160 A1) showed phospholipase and lipase co-activity of derived has additional outstanding properties. Thus, while CalB displayed the highest enzyme activity to the synthetic substrate tributyrin, the phospholipase/lipase MPlaG had, compared to CalB, superior activity to the natural substrate olive oil and the phospholipase substrate phosphatidylcholine that cannot hydrolyzed by CalB. The enzyme was hardly inhibited by polar organic solvents and is therefore a phospholipase/lipase fully usable in the organic solvent conditions. Additionally, based on the preliminary functional characteristics of optimal temperature and hydrolytic capacity, a novel thermophilic and thermostable lipase from a

Thailand hot spring metagenomic library was suggested to have properties similar to those of CalB (Tirawongsaroj et al. 2008); however, their performance for synthetic applications was not tested. This is of particular importance because the potential use of an esterase/lipase in biocatalysis and in chemical industry in general should take into consideration not only the hydrolytic performance but also the behavior under synthetic conditions.

Biocatalysis innovation guided by metagenomics

A next logical step in metagenomics research is to test the isolated enzymes in biotransformations with high application potential in the chemical industry. For this reason, it is a priority to identify and catalogue the chemical reactions of interest to industry, for which the enzymes already (and yet to be) discovered in metagenomes should be tested. We provided, in Table II and below, a few relevant examples that deserve investigation at an industrial scale.

In general terms, the new lipase and esterase enzymes that are needed should cover two main classes of chemical reactions: *i*) all reactions (both the synthesis and the hydrolysis) that currently cannot be solved properly at industrial settings because no enzyme that supports such transformations are known; and *ii*) those reactions that some already available promiscuous lipases and esterases have the ability to perform but have not been exploited commercially because of their low performance under industrial conditions.

The first group would include all chemical reactions involving carbon-oxygen bonds, such as the hydrolysis and synthesis of esters, epoxides and amides. Most of the LIP-EST enzymes can recognize the carboxyl group on fatty acids, alkanes and, in general terms, long chain structures where the carboxyl group is located at one end. However, the process does not work as well when the carboxyl group is near a bigger, more complex, and especially less flexible, structure, such as oleanolic acid, abietic acid, or generally speaking naphthene, phenanthrene or terpenoid

structures (Martínez-Martínez et al. 2014). Another problem to be addressed is how to perform the reaction when the substrate/substrates (the acid and/or the alcohol molecules) are nearly insoluble in water or in solvents where the LIP–EST enzymes are active. In such a case, the enzyme cannot easily access the substrates due to their low solubility under reaction conditions. In this context, lipase-based conversions involving abietic acid and complex resinic acids, such as pine/colofony, which are only soluble in trementine oil (Illanes et al. 2008; Aouf et al. 2014), and hydrolysates of collagen (or other similar substances) and polyols, such as pentaerithritol, which are almost insoluble in all types of solvents, are already a challenging issue. For all of these cases, several strategies have been examined: *i*) the use of water/solvent biphasic systems; *ii*) the use of substrate immobilization systems; *iii*) the use of ionic liquids; and *iv*) the generation of eutectic points between two solvents. However, unfortunately, it is clear that none of these strategies are currently possible at industrial scales at competitive costs, not only due to the complexity of the operation but also due to their low yield (Adlercreutz 2012; Hrydziouszko et al. 2014).

The second group would include all chemical reactions that involve carbon-carbon bonds, the opening/fission of aromatic rings through the epoxidation of the ring, the epoxidation of unsaturated carbon-carbon bonds, or carbon-carbon bond formation by the addition of 1,3-dicarbonyls to α/β -unsaturated carbonyl compounds. The epoxidation of aromatic rings is a common strategy that microbial strains use for the degradation of celluloses, lignins and tannins. This is a very slow process, even in nature, and the available LIP–EST enzymes can only act, for industrial purposes, upon small molecules (Aouf et al. 2013). Complex terpenoid or esterol structures are still not able to be epoxidized using EST–LIP enzymes for industrial purposes. The epoxidation of unsaturated carbon-carbon bonds is not an easy problem to solve when other carboxyl end groups are present in the substrate structure because this carboxyl end group interferes with the desired reaction. Thus, known lipases or esterases do not have enough

capacity for the epoxidation of unsaturated carbon-carbon bonds. The addition of 1,3-dicarbonyls to α/β -unsaturated carbonyl compounds is a very common organic synthesis reaction, known as Michael addition. Although, some lipases can perform this reaction (Svedendahl et al. 2005), its extension is very poor because it is highly dependent on the solubility of the substrates and the activation capacity of the enzymes in the solvent used.

Conclusion

Herein, the most recent literature and annotated sequences of esterases/lipases discovered using metagenome analyses are compiled and reviewed for first time. A special focus is given to the biochemical characterization and activities for processes of industrial interest. For that, we provide a detailed overview of the properties of 288 esterases/lipases with validated functions that have been isolated using a metagenome approach and (mostly partially) characterized. The most relevant properties include sequence homology, phylogenetic relationships, substrate range, selectivity, salt and solvent tolerance, and optimal temperature. This information revealed that the 288 validated EST-LIP enzymes from metagenome analyses reviewed herein are encoded by a high diversity of non-redundant sequences. It also provided an overview of the potential of such enzymes for use in chemical transformations. Thus, a comparative analysis revealed that approximately 60% of them have properties of interest in biocatalysis, such as tolerance to solvents, a wide range of temperatures and ample substrate specificity. Such sequences are identified to help future screening programs of similar sequences in metagenomes. We also emphasized that both known and yet to be discovered EST-LIP enzymes in metagenomes should be tested in unresolved biotransformations of interest in the chemical industry, some of which are herein suggested.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the European Community project MAMBA (FP7-KBBE-2008-226977), MAGIC-PAH (FP7-KBBE-2009-245226), ULIXES (FP7-KBBE-2010-266473), MicroB3 (FP7-OCEAN.2011-2-287589), KILL-SPILL (FP7-KBBE-2012-312139) and Royal Society UK-Russia Exchange Grant (IE130218). We thank EU Horizon 2020 Program for the support of the Project INMARE H2020-BG-2014-2634486. This work was further funded by grants BIO2011-25012, PCIN-2014-107 and BIO2014-54494-R from the Spanish Ministry of Economy and Competitiveness. The present investigation was funded by the Spanish Ministry of Economy and Competitiveness, the UK Biotechnology and Biological Sciences Research Council (BBSRC) and the German Federal Ministry of Education and Research (BMBF) within the ERA NET-IB2 program, grant number ERA-IB-14-030. The authors gratefully acknowledge the financial support provided by the European Regional Development Fund (ERDF).

Declaration of interest: The authors report no declarations of interest.

References

- Adlercreutz P. 2012. Immobilisation and application of lipases in organic media. *Chem Soc Rev* 42:6406-6436.
- Alcaide M, Stogios PJ, Lafraya A, Tchigvintsev A, Flick R, Bargiela R, Chernikova TN, Reva ON, Hai T, Leggewie CC, Katzke N, La Cono V, Matesanz R, Jebbar M, Jaeger KE, Yakimov MM, Yakunin AF, Golyshin PN, Golyshina OV, Savchenko A, Ferrer M. 2015a. Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats. *Environ Microbiol* 17:332-345.
- Alcaide M, Tchigvintsev A, Martinez-Martinez M, Popovic A, Reva ON, Lafraya A, Bargiela R, Nechitaylo TY, Matesanz R, Cambon-Bonavita MA, Jebbar M, Yakimov MM, Savchenko A, Golyshina OV, Yakunin AF, Golyshin PN, Ferrer M. 2015b. Identification and characterization of

- carboxyl esterases of gill chamber-associated microbiota in the deep-sea shrimp *Rimicaris exoculata* by using functional metagenomics, *Appl Environ Microbiol* 81:2125-2136.
- Alcaide M, Tornés J, Stogios Peter J, Xu X, Gertler C, Di Leo R, Bargiela R, Lafraya Á, Guazzaroni M-E, López-Cortés N, Chernikova TN, Golyshina OV, Nechitaylo TY, Plumeier I, Pieper DH, Yakimov MM, Savchenko A, Golyshin PN, Ferrer M. 2013. Single residues dictate the co-evolution of dual esterases: MCP hydrolases from the α/β hydrolase family. *Biochem J* 454:157-166.
- Alcantara A. 2002. Heptyl oleate synthesis as useful tool to discriminate between lipases, proteases and other hydrolases in crude preparations. *Enzyme Microb Tech* 31:283-288.
- Alcántara AR, Sánchez JM. 2010. Utilización de hidrolasas en la preparación de fármacos e intermedios homóquimicos. *An R Acad Nac Farm* 76:259-305.
- Alnoch RC, Martini VP, Glogauer A, Costa ACdS, Piovan L, Muller-Santos M, de Souza EM, de Oliveira Pedrosa F, Mitchell DA, Krieger N. 2015. Immobilization and characterization of a new regioselective and enantioselective lipase obtained from a metagenomic library. *PloS One* 10:e0114945.
- Anobom CD, Pinheiro AS, De-Andrade RA, Aguiéiras EC, Andrade GC, Moura MV, Almeida RV, Freire DM. 2014. From structure to catalysis: recent developments in the biotechnological applications of lipases. *Biomed Res Int* 2014:684506.
- Aouf C, Durand E, Lecomtre J, Figueroa-Espinoza MC, Dubreucq E, Fulcranda H. 2014. The use of lipases as biocatalyst for the epoxidation of fatty acids and phenolic compounds. *Green Chem* 16:1740-1754.
- Arpigny JL, Jaeger KE. 1999. Bacterial lipolytic enzymes: classification and properties. *Biochem J* 343:177-183.

- Bertram M, Hildebrandt P, Weiner DW, Patel JS, Bartnek F, Hitchman T, Bornscheuer UT. 2008. Characterization of lipases and esterases from metagenomes for lipid modification. *J Am Oil Chem Soc* 85:47-53.
- Bommarius AS. 2015. Biocatalysis: A Status Report. *Annu Rev Chem Biomol Eng* 6:319-345.
- Bornscheuer UT. 2002. Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol Rev* 26:73-81.
- Brady L, Brzozowski AM, Derewenda ZS, Dodson E, Dodson G, Tolley S, Turkenburg JP, Christiansen L, Huge-Jensen B, Norskov L, et al. 1990. A serine protease triad forms the catalytic centre of a triacylglycerol lipase. *Nature* 343:767-770.
- Chistoserdova, L. 2014 Is metagenomics resolving identification of functions in microbial communities? *Microb Biotechnol* 7:1-4.
- Chow J, Kovacic F, Dall Antonia Y, Krauss U, Fersini F, Schmeisser C, Lauinger B, Bongen P, Pietruszka J, Schmidt M, Menyes I, Bornscheuer UT, Eckstein M, Thum O, Liese A, Mueller-Dieckmann J, Jaeger KE, Streit WR. 2012. The metagenome-derived enzymes lips and lipt increase the diversity of known lipases. *PloS One* 7:e47665.
- Daiha Kde G, Angeli R, de Oliveira SD, Almeida RV. 2015. Are lipases still important biocatalysts? a study of scientific publications and patents for technological forecasting. *PLoS One* 10:e0131624.
- de Lourdes Moreno M, Pérez D, García MT, Mellado E. 2013. Halophilic bacteria as a source of novel hydrolytic enzymes. *Life (Basel)* 3:38-51.
- Dung PT, Trung TQ, Kim KH. 2009. Preparative Resolution of Etodolac enantiomers by preferential crystallization method. *Arch Pharm Res* 32:1425-1431.
- Ewis HE, Abdelal AT, Lu C-D. 2004. Molecular cloning and characterization of two thermostable carboxyl esterases from *Geobacillus stearothermophilus*. *Gene* 329:187-195.

- Fernández-Álvaro E, Kourist R, Winter J, Böttcher D, Liebeton K, Naumer C, Eck J, Leggewie C, Jaeger KE, Streit W, Bornscheuer UT. 2010. Enantioselective kinetic resolution of phenylalkyl carboxylic acids using metagenome-derived esterases. *Microb Biotechnol* 3:59-64.
- Fernández-Llorente G, Palomo JM, Mateo C, Guisan JM, Fernández-Lafuente R. 2004. Resolution of paroxetine precursor using different lipases: influence of the reaction conditions on the enantioselectivity of lipases. *Enzyme Microb Tech* 34:264-269.
- Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Martins dos Santos VAP, Yakimov MM, Timmis KN, Golyshin PN. 2005. Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem Biol* 12:895-904.
- Ferrer M, Martinez-Martinez M, Bargiela R, Streit WR, Golyshina OV, Golyshin PN. 2015. Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb Biotechnol*, in press (doi: 10.1111/1751-7915.12309).
- Fotheringham IG, Grinter N, Pantaleone DP, Senkpeil RF, Taylor PP. 1999. Engineering of a novel biochemical pathway for the biosynthesis of L-2-aminobutyric acid in *Escherichia coli* K12. *Bioorg Med Chem* 7:2209-2213.
- Frolow F, Harel M, Sussman JL, Mevarech M, Shoham M. 1996. Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe2S ferredoxin. *Nat Struct Biol* 3:452-458.
- Gaur R, Khare, SK. 2011. Solvent tolerant Pseudomonads as a source of novel lipases for applications in non-aqueous systems. *Biocatal Biotrans* 5:161-171.
- Gonzalo G, Brieva R, Sanchez VM, Bayod M, Gotor V. 2001. Enzymatic resolution of trans-4-(4'-fluorophenyl)-3-hydroxymethylpiperidines, key intermediates in the synthesis of (-)-Paroxetine. *J Org Chem* 66:8947-8953.

- Handrick R, Reinhardt S, Focarete ML, Scandola M, Adamus G, Kowalczyk M, Jendrosseck D. 2001. A new type of thermoalkalophilic hydrolase of *Paucimonas lemoignei* with high specificity for amorphous polyesters of short chain-length hydroxyalkanoic acids. *J Biol Chem* 276:36215-36224.
- Houde A, Kademi A, LeBlanc D. 2004. Lipases and their industrial applications. *Appl Biochem Biotech* 118:155-170.
- Hrydziusko Z, Dmytryk A, Majewska P, Szymańska K, Liesiene J, Jarzębski A, Bryjak J. 2014. Screening of lipase carriers for reactions in water, biphasic and pure organic solvent systems. *Acta Biochim Pol* 61:1-6.
- Illanes A, Álvarez L, Álvaro G. 2008. Chemoselective transesterification of wood sterols by lipases. *Revc Colomb Biotechnol* 10:17-35.
- Jochens H, Hesseler M, Stiba K, Padhi SK, Kazlauskas RJ, Bornscheuer UT. 2011. Protein engineering of α/β -hydrolase fold enzymes. *Chembiochem* 12:1508-1517.
- José C, Toledo MV, Briand LE. 2015. Enzymatic kinetic resolution of racemic ibuprofen: past, present and future. *Crit Rev Biotechnol* 3:1-13.
- Kapoor M, Gupta M. 2012. Lipase promiscuity and its biochemical applications. *Process Biochem* 47: 555-569.
- Kazlauskas RJ, Bornscheuer UT. 1998., Biotransformations with Lipases, in.: *Biotechnology-Series* (Rehm HJ, Reed G, Pühler A, Stadler PJW, Kelly DR; eds.), Vol.8a, 37-191. Weinheim: Wiley-VCH.
- Kilic L. 2004. Lipase-catalyzed esterification of glycerol and oleic acid. *J Am Oil Chem Soc* 81:281-284.
- Kim E-Y, Oh K-H, Lee M-H, Kang C-H, Oh T-K, Yoon J-H. 2009. Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. *Appl Environ Microbiol* 75:257-260.

- Kim Y, Park J, Kin M. 2011. Dynamic kinetic resolution of amines and amino acids by enzyme- metal co-catalysis. *ChemCatChem* 3:271-277.
- Kourist R, Hari Krishna S, Patel JS, Bartnek F, Hitchman TS, Weiner DP, Bornscheuer UT. 2007. Identification of a metagenome-derived esterase with high enantioselectivity in the kinetic resolution of arylaliphatic tertiary alcohols. *Org Biomol Chem* 5:3310-3313.
- Kourist R, Jochens H, Bartsch S, Kuipers R, Padhi SK, Gall M, Böttcher D, Joosten H-J, Bornscheuer UT. 2010. The α/β -hydrolase fold 3DM database (ABHDB) as a tool for protein engineering. *ChemBioChem* 11:1635-1643.
- Kube M, Chernikova TN, Al-Ramahi Y, Beloqui A, Lopez-Cortez N, Guazzaroni ME, Heipieper HJ, Klages S, Kotsyurbenko OR, Langer I, Nechitaylo TY, Lünsdorf H, Fernández M, Juárez S, Ciordia S, Singer A, Kagan O, Egorova O, Petit PA, Stogios P, Kim Y, Tchigvintsev A, Flick R, Denaro R, Genovese M, Albar JP, Reva ON, Martínez-Gomariz M, Tran H, Ferrer M, Savchenko A, Yakunin AF, Yakimov MM, Golyshina OV, Reinhardt R, Golyshin PN. 2013. Genome sequence and functional genomic analysis of the oil-degrading bacterium *Oleispira antarctica*. *Nat Commun* 4:2156
- Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Göker M, Parker CT, Amann R, Beck BJ, Chain PS, Chun J, Colwell RR, Danchin A, Dawyndt P, Dedeurwaerdere T, DeLong EF, Detter JC, De Vos P, Donohue TJ, Dong XZ, Ehrlich DS, Fraser C, Gibbs R, Gilbert J, Gilna P, Glöckner FO, Jansson JK, Keasling JD, Knight R, Labeda D, Lapidus A, Lee JS, Li WJ, Ma J, Markowitz V, Moore ER, Morrison M, Meyer F, Nelson KE, Ohkuma M, Ouzounis CA, Pace N, Parkhill J, Qin N, Rossello-Mora R, Sikorski J, Smith D, Sogin M, Stevens R, Stingl U, Suzuki K, Taylor D, Tiedje JM, Tindall B, Wagner M, Weinstock G, Weissenbach J, White O, Wang J, Zhang L, Zhou YG, Field D, Whitman WB, Garrity GM, Klenk HP. 2014. Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS Biol* 12:e1001920

- Lan D, Popowicz GM, Pavlidis IV, Zhou P, Bornscheuer UT, Wang Y. 2015. Conversion of a mono- and diacylglycerol lipase into a triacylglycerol lipase by protein engineering. *Chembiochem* 16:1431-1434.
- Lee M-H, Lee C-H, Oh T-K, Song JK, Yoon J-H. 2006. Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. *Appl Environ Microbiol* 72:7406-7409.
- Liu P, Wang Y-F, Ewis HE, Abdelal AT, Lu C-D, Harrison RW, Weber IT. 2004. Covalent reaction intermediate revealed in crystal structure of the *Geobacillus stearothermophilus* carboxylesterase Est30. *J Mol Biol* 342:551-561.
- Madern D, Ebel C, Zaccai G. 2000. Halophilic adaptations of enzymes. *Extremophiles* 4:91-98.
- Martínez-Martínez M, Alcaide M, Tchigvintsev A, Reva O, Polaina J, Bargiela R, Guazzaroni M-E, Chicote Á, Canet A, Valero F, Rico Eguizabal E, Guerrero Mdel C, Yakunin AF, Ferrer M. 2013. Biochemical diversity of carboxyl esterases and lipases from Lake arreo (spain): a metagenomic approach. *Appl Environ Microbiol* 79:3553-3562.
- Martínez-Martínez M, Lores I, Peña-García C, Bargiela R, Reyes-Duarte D, Guazzaroni ME, Peláez AI, Sánchez J, Ferrer M. 2014. Biochemical studies on a versatile esterase that is most catalytically active with polyaromatic esters. *Microb Biotechnol* 7:184-191.
- Mayumi D, Akutsu-Shigeno Y, Uchiyama H, Nomura N, Nakajima-Kambe T. 2008. Identification and characterization of novel poly(DL-lactic acid) depolymerases from metagenome. *Appl Microbiol Biotechnol* 79:743-750.
- Meilleur C, Hupé JF, Juteau P, Shareck F. 2009. Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. *J Ind Microbiol Biotechnol* 36:853-861.

- Moghaddam R, Salimon J, Jelas Haron MD, Jahangirian H, Shah Ismail MH, Hosseini S, Reyazi M. 2014. Lipase epoxidation optimizing for *Jatropha curcas* oil using perlauric acid. *Digest J Nanomat Biostruc* 9:1159-1169.
- Monti D, Gazak R, Marhol P, Biedermann D, Purchartova K, Fedrigo M, Riva S, Kren V. 2010. Enzymatic kinetic resolution of silybin diastereoisomers. *J Nat Prod* 73:613-619.
- Nagarajan S. 2012. New tools for exploring "old friends-microbial lipases". *Appl Biochem Biotechnol* 168:1163-1196.
- Placido A, Hai T, Ferrer M, Chernikova TN, Distaso M, Armstrong D, Yakunin AF, Toshchakov SV, Yakimov MM, Kublanov IV, Golyshina OV, Pesole G, Ceci LR, Golyshin PN. 2015. Diversity of hydrolases from hydrothermal vent sediments of the Levante Bay, Vulcano Island (Aeolian archipelago) identified by activity-based metagenomics and biochemical characterization of new esterases and an arabinopyranosidase. *Appl Microbiol Biotechnol*, in press.
- Plou FJ, Cruces MA, Ferrer M, Fuentes G, Pastor E, Bernabé M, Christensen M, Comelles F, Parra JL, Ballesteros A. 2002. Enzymatic acylation of di- and trisaccharides with fatty acids: choosing the appropriate enzyme, support and solvent. *J Biotechnol* 96:55-66.
- Rao L, Xue Y, Zheng Y, Lu JR, Ma Y. A 2013. A novel alkaliphilic bacillus esterase belongs to the 13(th) bacterial lipolytic enzyme family. *PloS One* 8:e60645.
- Rauwerdink A., Kazlauskas RJ. 2015. How the same core catalytic machinery catalyzes 17 different reactions: the serine-histidine-aspartate catalytic triad of α/β -hydrolase fold enzymes. *ACS Catal* 5:6153–6176.
- Reyes-Duarte D, Ferrer M, García-Arellano H. 2012. Functional-based screening methods for lipases esterases and phospholipases in metagenomic libraries. *Methods Mol Biol* 861:101-113.

- Sabín JG, Varas FM, González CP, Saldaña FJ, Rebolledo VF, Santamaría VG. 2008. Síntesis enzimática de derivados enantioméricamente enriquecidos de cis- y trans- ciclopentano 1,2 diaminas. Patent WO 2010007202 A1.
- Secundo F, Carrea G, Soregaroli C, Varinelli D, Morrone R. 2001. Activity of different *Candida antarctica* lipase B formulations in organic solvents. *Biotechnol Bioeng* 73:157-163.
- Sekhon BS. 2010. Enantioseparation of Chiral Drugs – An Overview. *Int Journal of PharmTech Res* 2:1584-1594.
- Shanghai Hu, Carlos Alberto Martinez, Junhua Tao, William Eugene Tully, Patrick Kelleher, Yves Dumond. 2010. Specific separation of S-Pregabalin using a lipase from *Thermomyces lanuginosus*. US patent USP 12/254336.
- Sharma S, Kanwar SS. 2014. Organic solvent tolerant lipases and applications. *ScientificWorldJournal* 2014:625258.
- Singh A, Kaplan DL. 2006. In vitro enzyme induced vinyl polymerization. *Adv Polym Sci* 194:211-224.
- Singh, B.K. 2010. Exploring microbial diversity for biotechnology: the way forward. *Trends Biotechnol* 28:111-116.
- Slotema WF, Sandoval G, Guieysse D, Straathof A J, Marty A. 2003. Economically pertinent continuous amide formation by direct-catalyzed amidation with ammonia. *Biotechnol Bioeng* 82:664-669.
- Spickermann D, Kara S, Barackov I, Hollmanns F, Schwaneberg U, Duenkelmanns P, Leggewie C. 2014. Alcohol dehydrogenase stabilization by additives under industrially relevant reaction conditions. *J Mol Catal B: Enzymatic* 103:24-28.
- Stecher H, Faber KJ. 1997. Biocatalytic deracemization techniques: dynamic resolutions and Stereoinversions. *Synthesis* 1997:1-16.

- Steele HL, Jaeger KE, Daniel R, Streit WR. 2009. Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microbiol Biotechnol* 16:25-37.
- Sulaiman S, Yamato S, Kanaya E, Kim JJ, Koga Y, Takano K, Kanaya S. 2012. Isolation of a novel cutinase homolog with polyethylene terephthalate-degrading activity from leaf-branch compost by using a metagenomic approach. *Appl Environ Microbiol* 78:1556-1562.
- Suplatov DA, Besenmatter W, Svedas VK, Svendsen A. 2012. Bioinformatic analysis of α/β -hydrolase fold enzymes reveals subfamily-specific positions responsible for discrimination of amidase and lipase activities. *Protein Eng Des Sel* 25:689-697.
- Svedendahl M, Hult K, Berglund P. 2005. Fast carbon-carbon bond formation by a promiscuous lipase. *J Am Chem Soc* 127:17988-17989.
- Tchigvintsev A, Tran H, Popovic A, Kovacic F, Brown G, Flick R, Hajighasemi M, Egorova O, Somody JC, Tchigvintsev D, Khusnutdinova A, Chernikova TN, Golyshina OV, Yakimov MM, Savchenko A, Golyshin PN, Jaeger KE, Yakunin AF. 2014. The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes. *Appl Microbiol Biotechnol* 99:2165-2178.
- Theodosiou E, Katsoura M, Loutrari H. 2009. Enzymatic preparation of acylated derivatives of silybin in organic and ionic liquid media and evaluation of their antitumor proliferative activity. *Biotrans* 27:161-169.
- Tirawongsaroj P, Sriprang R, Harnpicharnchai P, Thongaram T, Champreda V, Tanapongpipat S, Pootanakit K, Eurwilaichitr L. 2008. Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *J Biotechnol* 133:42-49.
- Troeschel SC, Thies S, Link O, Real CI, Knops K, Wilhelm S, Rosenau F, Jaeger KE. 2012. Novel broad host range shuttle vectors for expression in *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas putida*. *J Biotechnol* 161:71-79.

- Turner NJ, Truppo MD. 2013. Biocatalysis enters a new era. *Curr Opin Chem Biol* 17:212-214.
- Viader-Salvadó JM, Gallegos-López JA, Carreón-Treviño JG, Castillo-Galván M, Rojo-Domínguez A, Guerrero-Olazarán M. 2010. Design of thermostable beta-propeller phytases with activity over a broad range of pHs and their overproduction by *Pichia pastoris*. *Appl Environ Microbiol* 76:6423-6430.
- Winkler FK, D'Arcy A, Hunziker W. 1990. Structure of human pancreatic lipase. *Nature* 343:771-774.
- Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12:635-645.
- Yoon S, Kim S, Ryu Y, Kim TD. 2007. Identification and characterization of a novel (S)-ketoprofen-specific esterase. *Int J Biol Macromol* 41:1-7.
- Zheng J, Liu C, Liu L, Jin Q. 2013. Characterisation of a thermo-alkali-stable lipase from oil-contaminated soil using a metagenomic approach. *Syst Appl Microbiol* 36:197-204.
- Zisis T, Freddolino PL, Turunen P, van Teeseling MC, Rowan AE, Blank KG. 2015. Interfacial activation of *Candida antarctica* lipase b: combined evidence from experiment and simulation. *Biochemistry* 54:5969-5979.
- Zuhse R, Leggenwie C, Hollmann F, Kara S. 2015. Scaling-up of “smart cosubstrate” 1,4-butanediol promoted asymmetric reduction of ethyl-4,4,4-trifluoroacetoacetate in organic media. *Org Process Res Dev* 19:369-372.

Supplementary Table I. List of metagenomic sequences encoding EST–LIP proteins with validated functions described in the NCBI database. Reference (DOI), accession number of the corresponding sequence, the reported experimental activity assigned to the sequence, the thermal, solvent/salt and reactivity characteristics, the pI values, and the identity to homologous sequences in NCBI, are indicated. ND, not determined or not available.

Table I. Summary list of EST–LIP proteins with outstanding synthetic and hydrolytic activities.

Chemicals ¹	Accession number / reference ¹
Synthetic reactions	
Olive oil + methanol → methyl oleate	AY304500
Sucrose + vinyl laurate → mono-lauryl sucrose	SRA059294 (LAE6)
Triacetin + methanol → methyl acetate	EU660533.1
1-Propanol + lauric acid → 1-propyl laurate	AFS34517.1
1-Tetradecanol + myristic acid → 1-tetradecyl myristate	AFS34517.1
(R,S)-1-phenylethanol + vinyl acetate → (R)-1-phenylethyl acetate	KM023399
Hydrolytic reactions	
Poly(ε-caprolactone); polyethylene terephthalate	AEV21261.1; 4EB0
Poly-(DL-lactic acid)	KF801579; KC986400; KC986401; AB302136
3,7-Dimethyl-1,6-octadien-3-yl acetate	ACH88047.1
Plant-like polymers	CAJ19128.1; CAJ19130.1; CAJ19122.1; CAJ19113.1; CAJ19109.1; AEI54552.1; AEI54551.1; AKA87410.1; KC438416; KF709432; KF705200; HQ147564
Phospholipids	EU285670.1; ABN58716.1
Pyrethroid pesticides; Rho-nitrophenyl esters	ACJ07038.1
7-[3-Octylcarboxy-(3-hydroxy-3-methyl-butyloxy)]-coumarin	AAAY90130.1
1,2-Di-O-lauryl-rac-glycero-3-glutaric acid 6'-methylresorufin ester	AAZ67909.1
1-Acetyl, 3-acetyl and 1,3-diacetyl esters of BODIPY FL chloramphenicol	AEL88620.1
Chiral esters	AFS34517.1; ABI94371.1; 4P6B_A; AAY45707.1; AAY45707.1; AEH57833.1; AEH57832.1; KM023399; EU195806; AAY90130.1; AAZ48934.1; AAY90130.1; SRA059294; KF831416; KF831417; KC986402; KR919661; KP861227; KP861228; JX133669; JX133670; JX133672-JX133675; KF831414; KF831415; KF831419-KF831421; KF831416; KF831417; KC986402; ABI94371.1; Ferrer et al 2005.
Aromatic ring fission products (carbon-carbon bonds hydrolysis)	Alcaide et al. 2013
Lactones and epoxides	SRA059294; Alcaide et al 2015a

¹For extensive details see Supplementary Table I.

Table II. Some examples of esterase and lipase-based biotransformations of interest in the chemical industries.

Type of reaction	Reference
Separation of enantiomers in a racemic mixture	
Ibuprofen	José et al. 2015
Dexketoprofen	Alcántara & Sánchez 2010
Phytosterol	Illanes et al. 2008
Sibilinin	Theodosiou et al. 2009; Monti et al. 2010
Carbinol	Gonzalo et al. 2001; Fernández-Llorente et al. 2004
alfa-Aminobutyric acid	Stecker & Faber, 1997
Etodolac	Dung et al. 2009; Sekhon 2010
1-Benzoyl- deca-hydroquinoline	Fotheringham et al. 1999
1-Amino 2-propanol	Sabín et al. 2008
Pregabalin	Shanghui et al. 2010
D-Lysine	Kim et al. 2010
Esterification and transesterification reactions	
Abietic acid and glycerin or trimethylolpropane	Illanes et al. 2008; Aouf et al. 2014
Polyacrylic acid and stearic alcohol	Secundo 2001; Singh & Kaplan 2006
Triglycerides and glycerin	Kilic 2004
Triglycerides	Alcántara 2002
Hydrolysis reactions	
Production of fatty acids from triglycerides	Houde et al. 2004
Other reactions of interest	
Epoxidation of triglycerides	Aouf et al. 2014; Moghaddam et al. 2014
Amidation of collagen and fatty acids	Slotema et al. 2003

Figure legends

Figure 1. The unrooted circular neighbor-joining tree indicating phylogenetic positions of polypeptide sequences of metagenomics EST–LIP like proteins with validated functions. GenBank® and PDB accession numbers are indicated. Lipase/esterase families are depicted according to Arpigny and Jaeger (1999) and further classifications (Rao et al. 2013). Abbreviations, as follows: GS-F, Genome Sequences assigned to an esterase/lipase Family (in bold letters); MGS, MetaGenomic protein Sequences. Sequences encoding EST–LIP from the α/β -hydrolase fold superfamily which have been reported to be solvent tolerant (blue), salt-tolerant (red) and showing outstanding reactivity (indicated with a black circle) are specifically indicated. A specific group of sequences that encode EST–LIP enzymes, which are phylogenetically related but with quite distinct reactivity are specifically indicated; this group includes hydrolases able to hydrolyse (S)-menthylacetate, or (S)-methyl 3-phenylglycidate or (S)-ketoprofen ester or poly-(DL-lactic acid) or β -lactam.

Figure 2. Pairwise identities within all sequences encoding esterase-lipases discovered in metagenomes. This figure compiles the analysis of 82944 comparisons that correspond to the pairwise identities when each of the sequences was compared with the rest of the 288 sequences ($288 \times 288 = 82944$ datasets in total). The segments of pairwise identity values are represented on the Y axis, while the X axis represents the frequency of those values within all pairwise comparisons.

Figure 3. Number of EST–LIP discovered in metagenomes for which salt- and solvent-tolerance has been experimentally reported. As shown, activity in the presence of salt-solvent has not been determined (ND) for 210 out of 288 EST–LIP reported. For details see Supplementary Table I.

Figure 4. Results of pI calculations of sequences encoding esterase-lipases discovered in metagenomes. The segments of pI values are represented in the Y axis, while the X axis represents the number of proteins within all pI segments. The pI can be used as an estimation of the presence of acidic amino acids. For details see Supplementary Table I.

Figure 5. Optimal temperature for activity for the 168 EST-LIP discovered in metagenomes for which this value was reported. The segments of optimal temperature values are represented in the Y axis, while the X axis represents the number of proteins within all thermal segments. For details see Supplementary Table I.

Figure 1

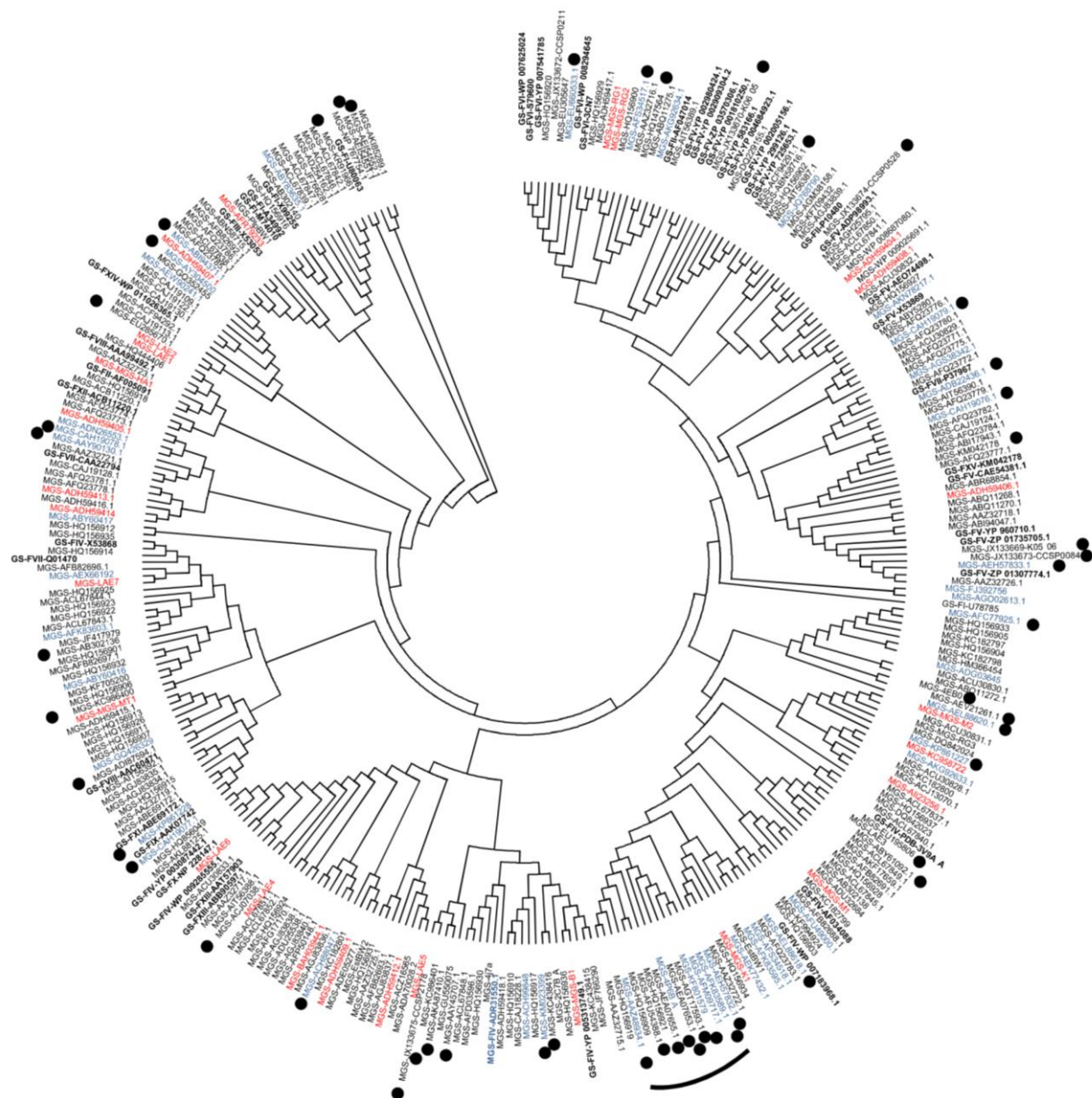


Figure 2

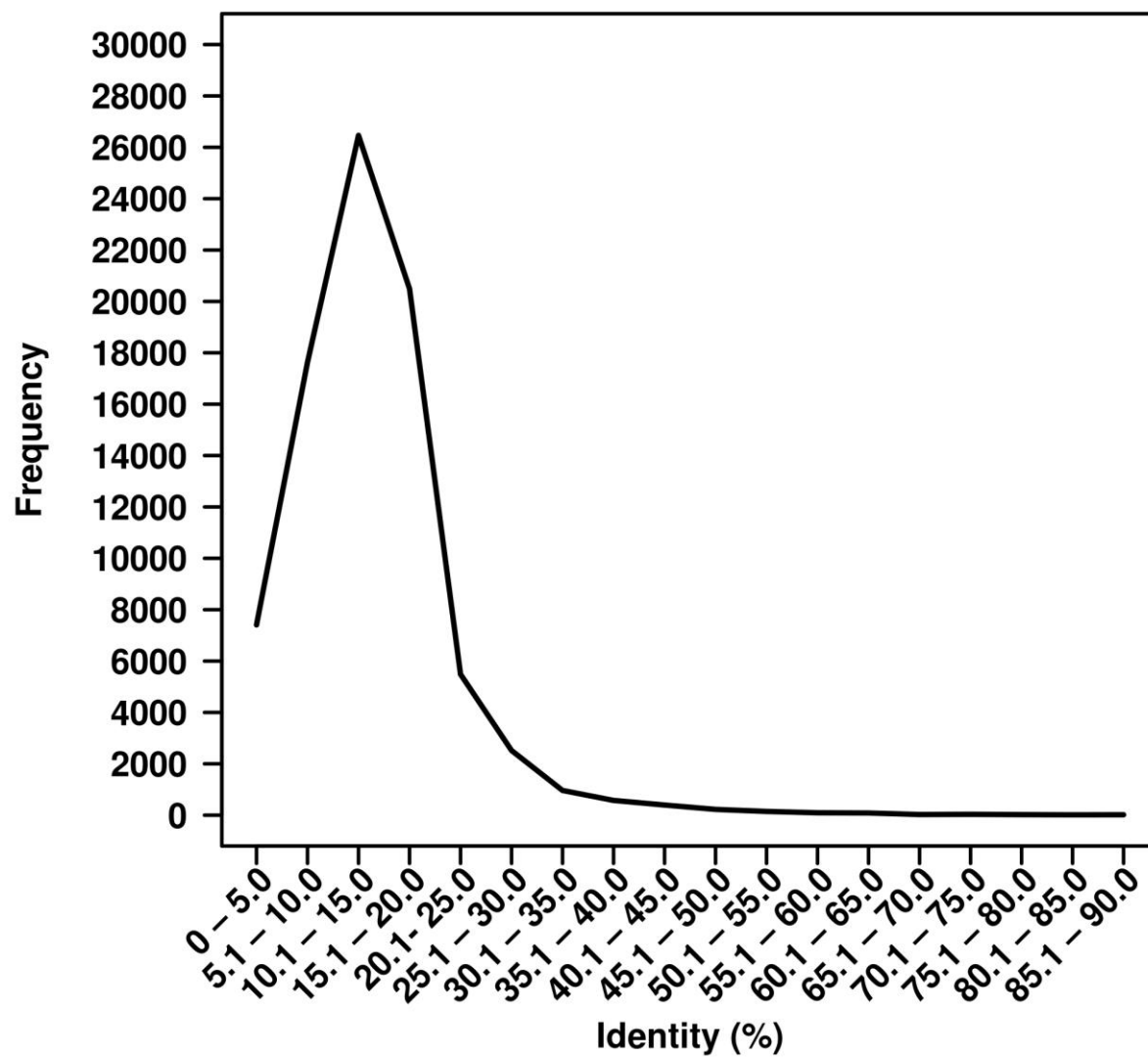


Figure 3

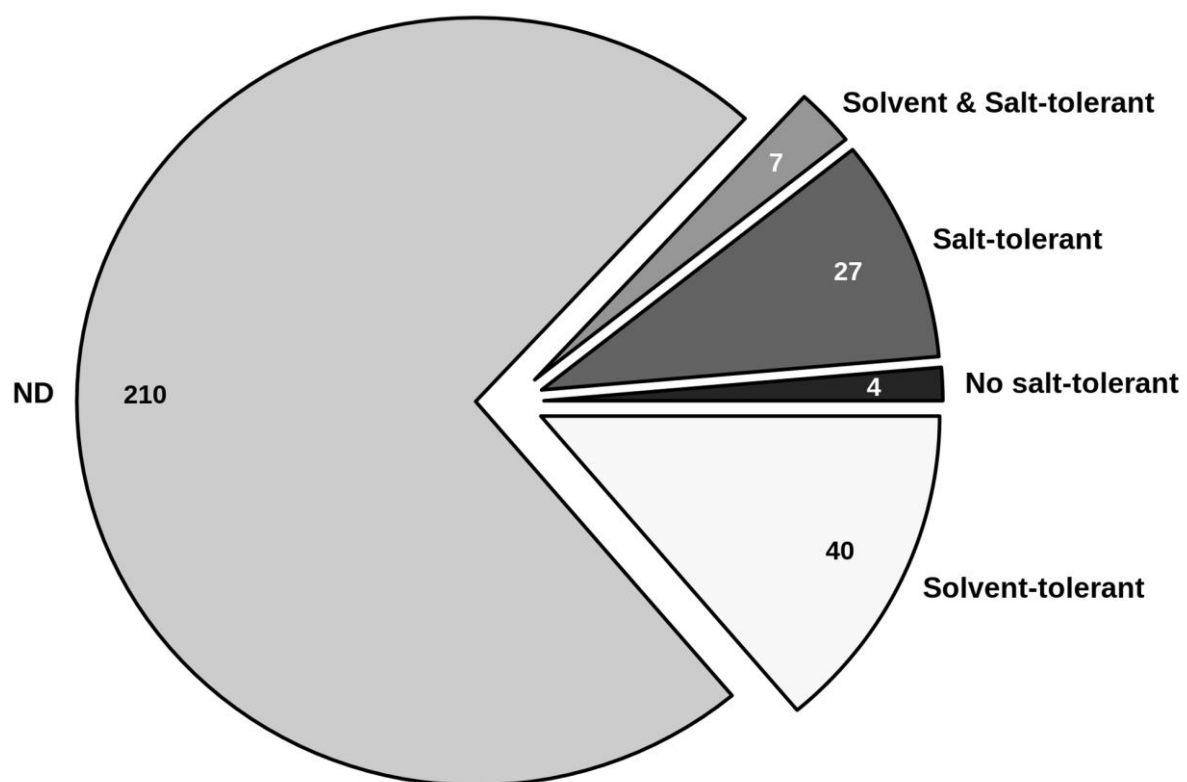


Figure 4

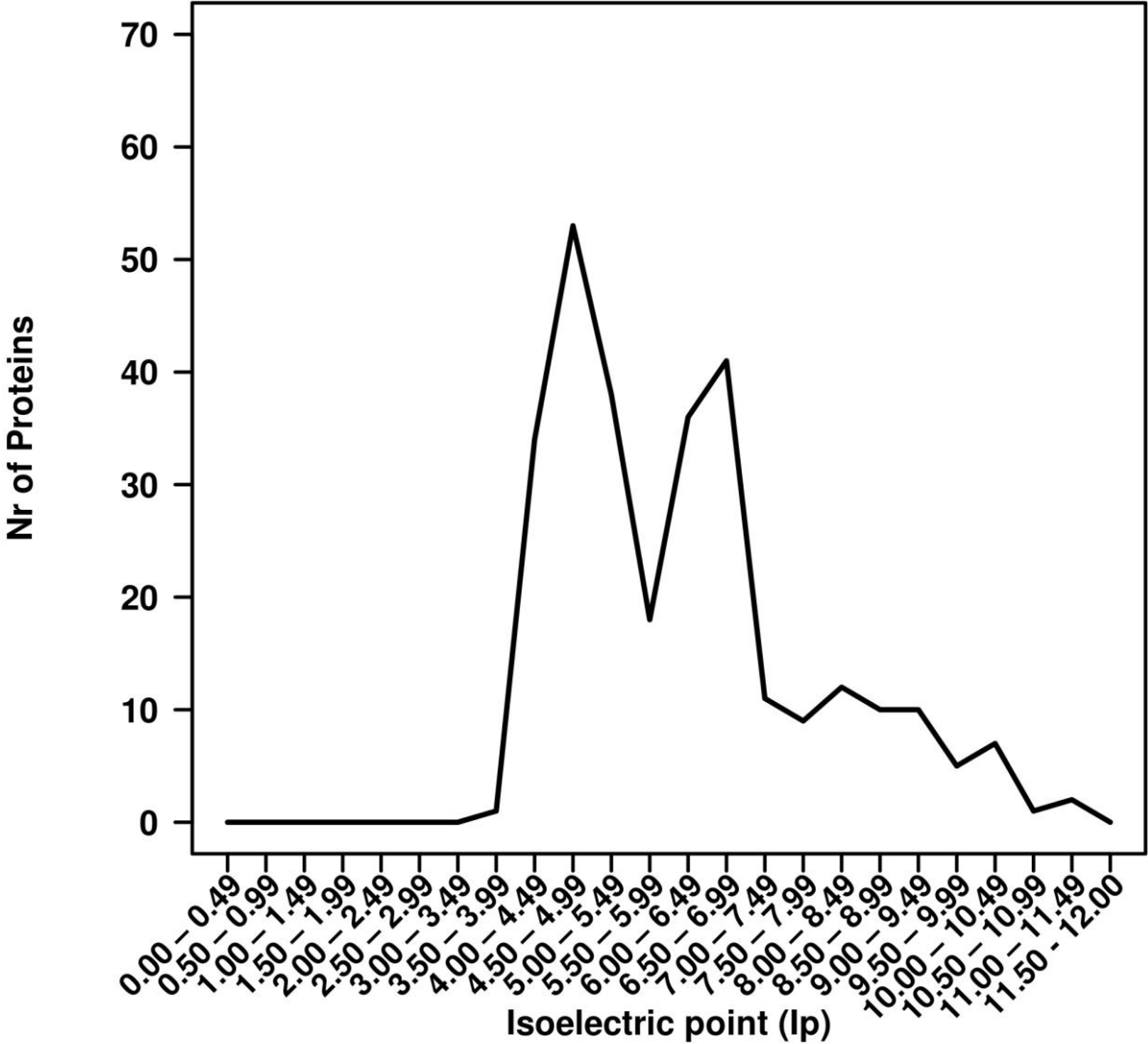


Figure 5

